

## RESEARCH ARTICLE

## Open Access

# Mitochondrial genome deletions and minicircles are common in lice (Insecta: Phthiraptera)

Stephen L Cameron<sup>1,2\*</sup>, Kazunori Yoshizawa<sup>3</sup>, Atsushi Mizukoshi<sup>3</sup>, Michael F Whiting<sup>4</sup> and Kevin P Johnson<sup>5</sup>**Abstract**

**Background:** The gene composition, gene order and structure of the mitochondrial genome are remarkably stable across bilaterian animals. Lice (Insecta: Phthiraptera) are a major exception to this genomic stability in that the canonical single chromosome with 37 genes found in almost all other bilaterians has been lost in multiple lineages in favour of multiple, minicircular chromosomes with less than 37 genes on each chromosome.

**Results:** Minicircular mt genomes are found in six of the ten louse species examined to date and three types of minicircles were identified: heteroplasmic minicircles which coexist with full sized mt genomes (type 1); multigene chromosomes with short, simple control regions, we infer that the genome consists of several such chromosomes (type 2); and multiple, single to three gene chromosomes with large, complex control regions (type 3). Mapping minicircle types onto a phylogenetic tree of lice fails to show a pattern of their occurrence consistent with an evolutionary series of minicircle types. Analysis of the nuclear-encoded, mitochondrially-targeted genes inferred from the body louse, *Pediculus*, suggests that the loss of mitochondrial single-stranded binding protein (mtSSB) may be responsible for the presence of minicircles in at least species with the most derived type 3 minicircles (*Pediculus*, *Damalinia*).

**Conclusions:** Minicircular mt genomes are common in lice and appear to have arisen multiple times within the group. Life history adaptive explanations which attribute minicircular mt genomes in lice to the adoption of blood-feeding in the Anoplura are not supported by this expanded data set as minicircles are found in multiple non-blood feeding louse groups but are not found in the blood-feeding genus *Heterodoxus*. In contrast, a mechanistic explanation based on the loss of mtSSB suggests that minicircles may be selectively favoured due to the incapacity of the mt replisome to synthesize long replicative products without mtSSB and thus the loss of this gene lead to the formation of minicircles in lice.

**Background**

The gene content, order and structure of bilaterian mitochondrial (mt) genomes are amongst the most stable genomic systems known. Gene content within bilaterians, with few exceptions, consists of 37 genes, 13 protein coding genes (PCGs), two ribosomal RNAs (rRNAs) and 22 transfer RNAs (tRNAs). This gene content is stable across 32 phyla, representing 78% of life and 555 million years of evolution. The exceptions such as the loss of the smallest PCG in higher nematodes [1] and of tRNAs in some mites (e.g. [2]), affect only genes which are physically small and these losses are confined to groups

representing a small portion of the bilaterian diversity [3]. Gene order is variable but in several instances an ancestral arrangement is shared by the majority of species in enormous groups such as the Superphylum Ecdysozoa (arthropods, velvet worms and priapulids) and for the chordates including most vertebrates [4-7]. Again exceptions such as the highly rearranged mt genomes of lice [8] and wasps [9] within arthropods or those of plethodontid salamanders within vertebrates [10] are comparatively rare and confined to relatively derived portions of each group. Mt genome structure within bilateria is almost universally a single circular genome although it is more variable in lower metazoa and protists [11-13]. This great stability of gene content, arrangement and structure across such an enormous age and diversity of life suggests strong stabilizing selection for an optimal

\* Correspondence: [sl.cameron@qut.edu.au](mailto:sl.cameron@qut.edu.au)<sup>1</sup>Discipline of Biogeosciences, Faculty of Science & Technology, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia  
Full list of author information is available at the end of the article

bilaterian mt genome. The few animal groups which diverge from this pattern are thus of particular utility for investigating the potential causes of this near-universal genomic stability.

Recently, two studies have independently reported the most marked departures from this general pattern of mt genomic stability within bilateria found to date: the break up of the canonical single chromosome mt genome into multi-chromosome, minicircular mt genomes. Gibson et al. [14,15] reported that the mt genome of the root parasitic nematode *Globodera* and near relatives was spread over six minicircular chromosomes. Each chromosome had between 2 and 19 genes, with several of the PCGs occurring on multiple chromosomes with evidence for modest nucleotide divergences between these different chromosome copies. Collectively the canonical set of 37 mt genes was present although no single chromosome contained all of the genes. Thus Gibson et al. [14] described the mt genome of *Globodera* to be a mosaic spread across the different chromosomes. A more extreme example was reported by Shao et al. [16] from the human body louse, *Pediculus* in which the mt genome was spread over at least 18 minicircular chromosomes, each containing one to three genes plus a large non-coding region containing sequence blocks that were highly conserved between different chromosomes. Again the canonical 37 mt genes were collectively present across the set of 18 mt chromosomes; however there was no evidence for a full-sized chromosome which contained all 37 genes in *Pediculus* or its close relatives.

The finding of multi-chromosomal, minicircular mt genomes in these groups is even more remarkable when compared against our growing knowledge of mt diseases in humans [17]. A wide range of human diseases have been found to result from the deletion of portions of the mt genome a state termed  $\Delta$ mtDNA.  $\Delta$ mtDNA is caused by either mutations within the mt genome itself or defective versions of the nuclear encoded mtgenome maintenance genes [18-21]. The progressive loss of mt function due to  $\Delta$ mtDNA mutations has also been hypothesized as the ultimate cause of aging: the mitochondrial theory of aging [22,23]. However, whether these mutations are sufficiently severe to cause the range of aging effects has been questioned [24]. Models of human mt disorders have been identified from screens of *Drosophila* mutants lacking or differing in the same replication and nucleotide transport proteins [25], highlighting the fact that these systems have been maintained over large evolutionary distances and are likely shared across bilaterians. The replacement of the canonical mt genome structure in *Pediculus* and *Globodera* with a putatively "pathological" structure as their standard suggests a major departure from these common mt genome maintenance pathways. Accordingly, further

investigation of mt genome evolution within these groups could well shed light on the forces shaping stabilizing selection on animal mt genome structure in general.

The mt genomes of lice (Insecta: Psocodea: Phthiraptera) are known to be highly rearranged [8,26,27] and the finding of multi-chromosomal, minicircular mt genomes within *Pediculus* thus must be considered within this context. Recent phylogenetic analyses of lice suggest that the sucking lice (Suborder Anoplura), of which *Pediculus* is a representative, may be derived from within the largest of the three chewing louse suborders, the Ischnocera [28-31]. To date mt genomes are available for just two ischnoceran lice, *Campanulotes* [27] and *Bothriometopus* [8] and neither study reports evidence for multi-chromosomal or minicircular genomes. This study thus had three aims:

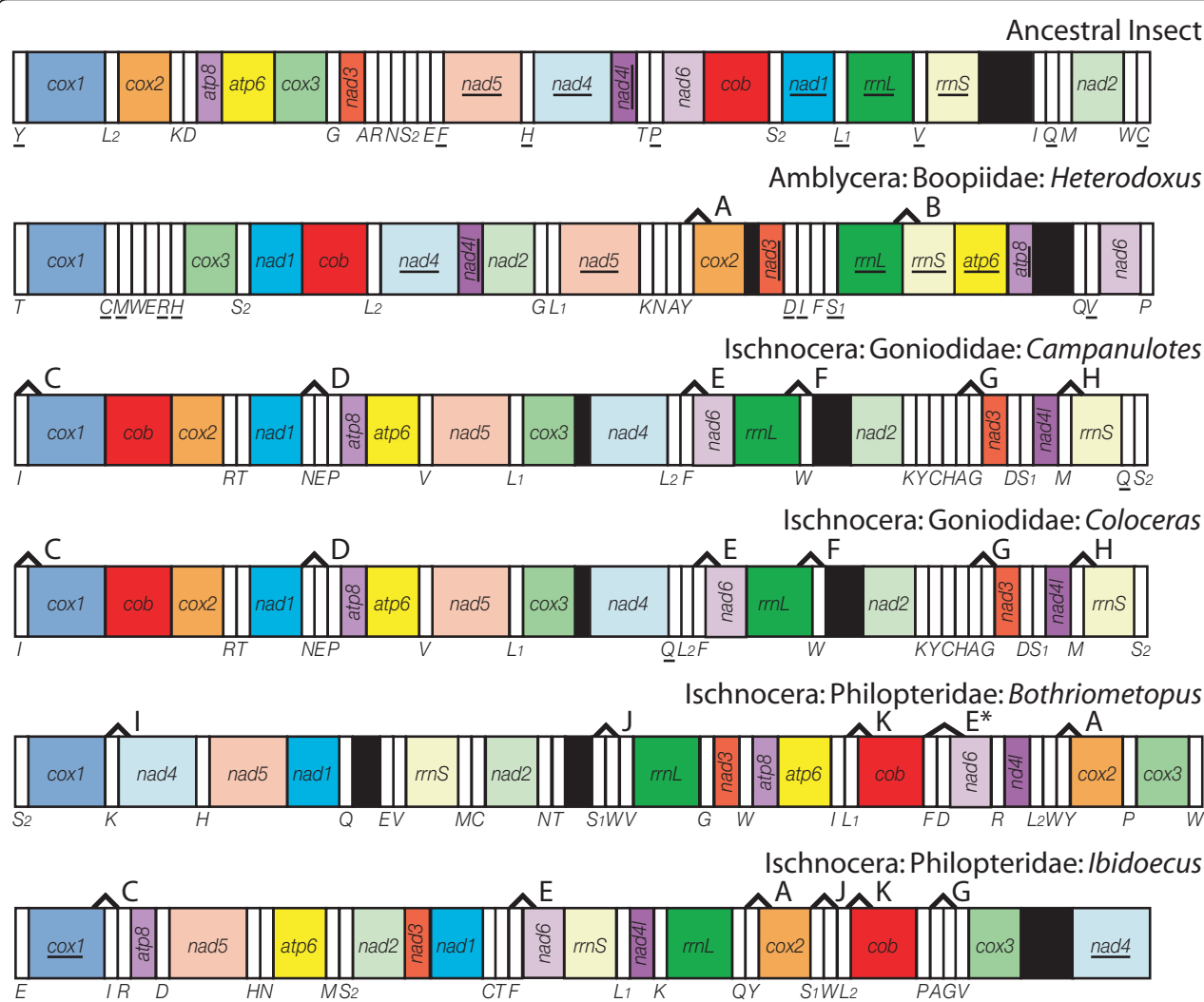
- 1) conduct a broader survey of ischnoceran lice for the presence of minicircular mt genomes;
- 2) compare genome structure to the evolutionary history of lice; and,
- 3) mine the recently published nuclear genome of *Pediculus* for nuclear-encoded, mitochondrially-targeted (ne-mt) genes to test if intrinsic factors in their genetic background might be responsible for the minicircularisation of louse mt genomes.

## Results

### Mini-circles are common in lice

Mt genomes from representatives of six ischnoceran genera were sequenced for this study representing the three most widely accepted ischnoceran families: Philopteridae (*Ibidoecus*, *Anaticola*, *Philopterus* and *Quadriceps*), Goniodidae (*Coloceras*) and Trichodectidae (*Damalinia*). Complete mt genome sequences including all protein-coding, rRNA and tRNA genes were determined for *Ibidoecus* (14,908 bp, GenBank accession number: JN122005) and *Coloceras* (14,868 bp; JN122000) (Figure 1). Multigene, minicircular genomes including only a subset of the canonical 37 mt genes were sequenced from *Anaticola* (8118 bp, 18 full length genes; JN121999), *Philopterus* (3721 bp, six genes; JN122006), *Quadriceps* (2553bp, six genes; JN21998) and *Coloceras* (7649 bp, 22 full and two partial genes; JN122001) in addition to the full sized mt genome found in *Coloceras* (Figure 2). Three minicircular genomes each including only a single protein coding or rRNA gene plus one tRNA gene were found in *Damalinia* (2079, 2083, 2306 bp; JN122002, JN122003, JN122004) (Figure 2). Complete details of the sequencing strategies followed to sequence each mt genome are provided in Additional File 1.

The minicircular genomes found in this study fall into three different categories. The *Coloceras* minicircle is one of two heteroplasmic genome types present in this



**Figure 1 Gene maps of complete louse mt genomes.** Circular genomes have been arbitrarily linearised on the tRNA gene immediately upstream of *cox1* to allow ease of comparison. tRNA genes are designated by the single letter amino acid code. Genes which are underlined are encoded on the opposite strand from the majority of genes in that genome. Black gene blocks represent putative control regions. Shared, derived gene boundaries are labeled above each genome; A: *trnY-cox2*; B: *rnsL-rnsS*; C: *trnI-cox1*; D: *trnN-trnE*; E: *trnF-nad6*; E\*: *trnF-trnX-nad6*; F: *rnsL-trnW*; G: *trnA-trnG*; H: *trnM-rrnS*; I: *trnK-nad4*; J: *trnS-trnW*; K: *trnL-cob*.

individual louse and has been generated by the deletion of approximately half the mt genome (7218bp covering seven tRNAs, six full length PCGs and portions of two additional PCGs). Heteroplasmy here is used to describe the existence of multiple genome types, not sequence divergences between homologous genes found on different chromosomes as this was not observed in *Coloceras*. The deletion occurs between two poly-T regions, a stretch of 28 T's from bases 710 to 737 in *cob* and of 19 T's from bases 1256 to 1274 in *nad4*. Similar deletions within protein coding genes were detected from *Pediculus* due to recombination between regions of sequence homology however the recombination sites were not homopolymer stretches [32]. The deletion in the *Coloceras* mt genome was

consistently detected by amplifications spanning the deleted region; multiple independent primer pairs producing the same result with the shortest using primers 325 bp upstream and 73 bp downstream respectively of the deletion producing a 398 bp product in place of the expected 7600 bp product. Long poly-T stretches are common in the *Coloceras* mt genome. In addition to the two poly-T stretches associated with the deletion there are 23 other poly-T stretches of 10+ bases occurring in eight of the 13 PCGs, in *trnT* and in the major non-coding region (= putative control region). Such long mononucleotide repeats are uncommon in the coding regions of other insects where they are largely confined to the control region. Given this large number of repeats between which



deletions could potentially occur, it is perhaps surprising that only a single minicircle type was found. That this single type was found from multiple different PCR primer combinations with varying conditions also argues against it being a PCR artifact caused by polymerase slippage between poly-T regions.

A second minicircle type is found in *Anaticola*, *Philopterus* and *Quadriceps*. In each of these species we were unable to find a full sized mt genome which included all 37 genes. Unlike in the *Coloceras* minicircle, all genes found on these genomes are full length and there are no mononucleotide stretches of more

than eight bases. *Anaticola* and *Philopterus* each have a large non-coding region, 232 and 138 bp respectively, which may function as a control region however the identification of such regions in lice has been problematic as they lack the conserved regions found in most other insects. It thus seems likely that these minicircles are not the result of recent, in evolutionary timescales, deletions from a full sized mt genome as in *Coloceras*. We were able to amplify two additional mt genes from two of these species, *Philopterus* and *Quadriceps*, but unable to link them by long PCR to any of the genes on the minicircles suggesting that the mt genome of these species is now functionally multi-chromosomal. We have, however, not been able to successfully amplify the mt chromosomes on which these genes occur and so we can only infer, rather than demonstrate, a multi-chromosomal mt genome for these species. Multiple primer combinations were used to attempt to amplify the missing genes from each of the species both within individual genes by short PCRs and between genes by long PCRs however we were not able to amplify additional chromosomes from these three species. Next-generation sequencing methods which are not reliant on PCR amplification hold promise in completing these genomes; however, we were not in a position to attempt this as part of the current study. Indeed, next-gen sequencing was the key to completing the sequencing of the multi-partite *Pediculus* mt genome [16] as prior to the nuclear genome sequencing project [33], only two mt genes had been successfully sequenced from this species [27].

The third minicircle type was found in *Damalinia* and is very similar to that previously reported from *Pediculus* by Shao et al. [16]. Each of the three minicircles found in *Damalinia* consists of a single PCG or rRNA, a single tRNA and a large non-coding region which displays shared motifs between copies on different minicircles. The non-coding regions are 694bp (*cox1* containing minicircle), 846bp (*rrnL* containing minicircle) and 1089bp (*rrnS* containing minicircle) long; most of the length differences are due to macrorepeats which vary in number between different chromosomes. The non-coding portion of the genome consists of conserved regions broken up by macrorepeats (Figure 3). The first conserved region, immediately downstream of the coding regions, consists of approx. 150 bp and includes a large hair-pin loop which may function as an origin of replication. This is followed by a less conserved region of 250 bp which is repeated in the *rrnS* minicircle. The third region is another conserved sequence block 142 bp long in the *rrnS* minicircle but longer in the *rrnL* and *cox1* minicircles due to the presence of two different repeats 53 and 67bp long respectively. The fourth region of 75 bp in the *cox1* minicircle is also conserved

however there are additional highly divergent regions of 171 and 220 bp each in the *rrnL* and *rrnS* minicircles respectively. The overall minicircle structure in *Damalinia*, a small number of genes plus a very large non-coding region composed of regions highly conserved between different minicircles, is very similar to that previously reported for *Pediculus*. If the remainder of the *Damalinia* mt genome has the same structure there could be 12 or more additional minicircles each with a single protein-coding gene and one or more tRNAs.

### Genome Annotations & Evolutionary Trends

Gene order maps for the six ischnoceran species in this study are included in Figures 1 & 2. Gene order within the full sized *Coloceras* mt genome is almost identical to that of *Campanulotes* in that they differ by a single tRNA translocation of *trnQ*; these are the first two louse genera to be examined which share more than a handful of plesiomorphic gene boundaries that are common across all insect orders. Additionally, *Coloceras* and *Campanulotes* share the derived arrangement *trnN-trnE* with *Pediculus* but no other louse genera. *Ibidoecus* shares none of the plesiomorphic gene boundaries found in other insects and lacks even the four identified in other lice: *atp8-atp6* found in *Heterodoxus*, *Campanulotes*, *Coloceras* and *Bothriometopus*; *trnG-nad3* found in *Campanulotes*, *Coloceras* and *Bothriometopus*; *nad4-trnH-nad5* found in *Bothriometopus*; and *nad4L-nad4* found in *Heterodoxus*. In contrast *Ibidoecus* shares several derived gene boundaries with other lice: *trnY-cox2* is shared with *Heterodoxus*, *Bothriometopus* and *Pediculus*; *trnA-trnG* are shared with *Campanulotes* and *Coloceras*; *trnF-nad6* is shared with *Campanulotes*, *Coloceras* and *Pediculus* and with slight modification with *Bothriometopus* (*trnF-trnD-nad6*) and *Anaticola* (*trnF-trnG-nad6*); *trnS-trnW* is shared with *Bothriometopus* and *Pediculus*; *trnL-cob* are shared with *Bothriometopus* and *trnI-cox1* is shared with all ischnoceran lice except *Bothriometopus*.

The number of missing genes creates some uncertainty about the rearrangement patterns within the partial, minicircular mt genomes. However, none of the four species included here possess any of the plesiomorphic gene boundaries found across insects. Three of these species, *Anaticola*, *Philopterus*, and *Damalinia*, share the widespread derived gene boundary *trnI-cox1*. *Anaticola* shares the derived boundary *trnK-nad4* with *Bothriometopus* and *Pediculus* although it has three copies of *trnK* and it is unclear which of these copies is homologous and which are duplicates. *Philopterus* shares *trnM-rrnS* and *rrnL-trnW* with *Campanulotes*, *Coloceras* and *Quadriceps* and *rrnS-rrnL* with *Quadriceps* and *Heterodoxus*. This rearrangement is, however, sufficiently similar to the insect plesiomorphic arrangement *rrnS-trnV-rrnL* as

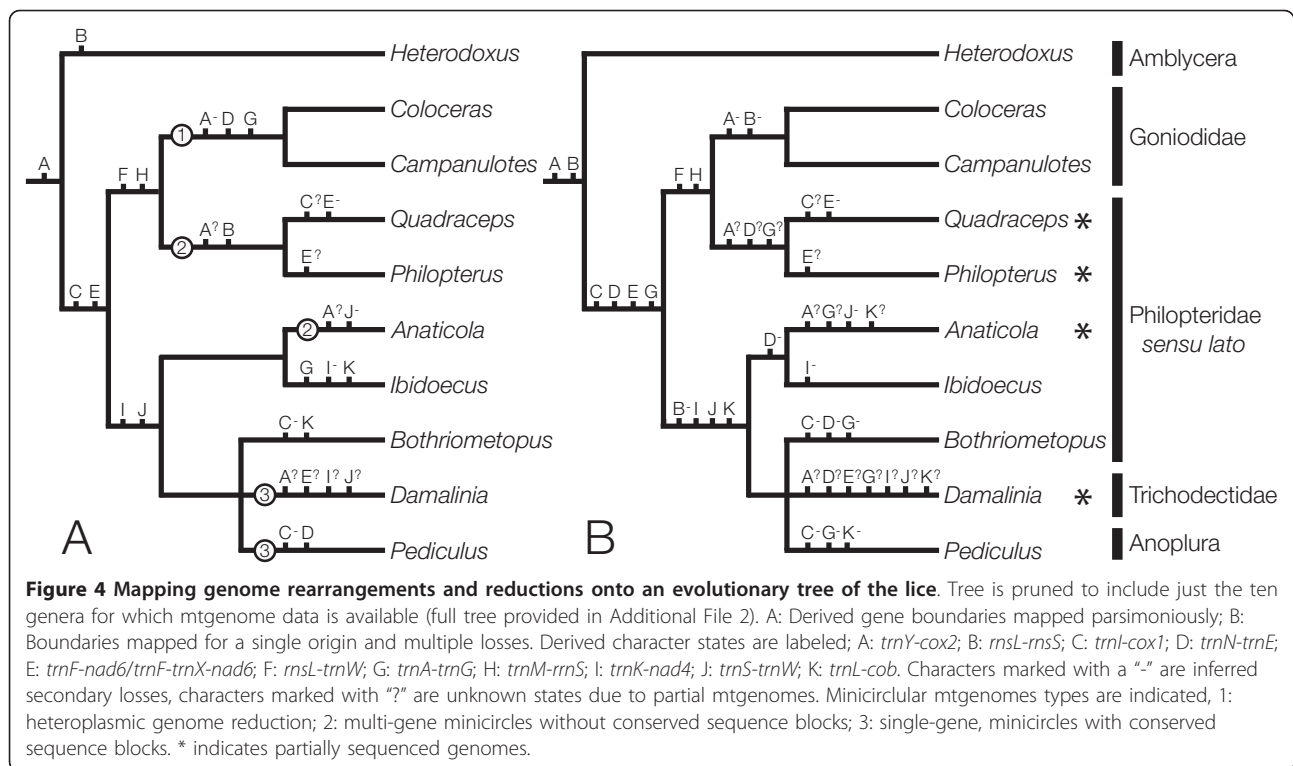




to be conceivably the product of convergent translocations of *trnV* out of this position occurring independently in each lineage.

Shared, derived gene boundaries were mapped onto a phylogenetic tree of the louse genera for which mtgenome data is available, either parsimoniously (Figure 4A), i.e. minimum number of inferred gene translocations even if that infers multiple independent origins of a given gene boundary, or with a single origin for each derived boundary and multiple losses (Figure 4B) (Figure 4 is a pruned copy of the tree in Additional file 2, and is reduced to just taxa for which mt genome data is available). The very closely related goniodid genera *Coloceras* and *Campanulotes* which have almost identical mt genome arrangements share 34 derived boundaries, however 27 of these are not present in any other genus and aren't considered further. There are 11 derived gene boundaries which are shared by at least two louse genera or the

goniodids plus at least one other louse (marked A - K, Figures 1, 2 & 4). Two of these, *rrnL-trnW* (derived gene boundary F) and *trnM-rrnS* (boundary H) are synapomorphic for the clade ((*Coloceras* + *Campanulotes*) + (*Quadriceps* + *Philoaterus*)) and are present in all members of that clade without secondary modification or loss. Five derived boundaries (A, C, E, I, J) are interpreted as synapomorphic for various clades under both character mapping schemes, but have been secondarily lost or are unknown in some members of that clade because the partial genomes available for some genera lack both genes. Another four derived boundaries (B, D, G, K) are most parsimoniously interpreted as having evolved convergently in two different genera or clades. However, a single origin optimization for these boundaries requires respectively, one loss and two unknown states (K), or two independent losses (B, D and G, with zero, two and three unknown states respectively).



Mapping genome reductions onto the same tree (Figure 4B) suggests that there is limited evolutionary pattern to the presence of mtgenome minicircles in different louse groups. Type 1 minicircles, heteroplasmic genome reductions from a full sized mtgenome, are confined to the Goniodidae. Type 2 minicircular mtgenomes were found in two different lineages, *Quadriceps*+*Philopterus* and *Anaticola*. Due to the polytomy between *Bothriometopus*, *Damalinia* and *Pediculus*, type 3 multichromosomal mtgenomes were not inferred to have a single evolutionary origin despite previous louse phylogenies strongly supporting the monophyly of *Damalinia*+*Pediculus* [31] which would suggest a single origin for type 3 minicircles.

#### Nuclear Encoded - Mitochondria Targeted Genes of Lice

We initially conducted an annotation of the *Pediculus humanus* ne-mt genes for the Human Body Louse Genome consortium [33] but a fuller discussion of these results follows here. The set of nuclear-encoded, mitochondrially targeted (ne-mt) genes identified from *Drosophila* consists of 315 genes. Reciprocal best blast matches were identified in the *Pediculus* nuclear genome for 295 (93.6%) of these genes with 68 of the 74 (91.9%) oxydative phosphorylation (OXPHOS) genes, 86/87 (98.9%) other metabolic genes, 108/113 (95.6%) protein synthesis genes, 12/15 (80%) DNA synthesis genes, 19/22 (86.4%) transport and cell rescue genes and 2/4 (50%) of the miscellaneous function genes. A table of the annotated genes, their blast

and alignment similarity statistics are included in Additional File 3. The numbers of OXPHOS [34-37] and protein synthesis homologues [38] identified from *Pediculus* are comparable to those identified in other insect genome projects. These two subsets are the best studied of the ne-mt genes and represent 59.4% of the mitochondrial gene set. In contrast the DNA synthesis set of *Pediculus* apparently lacks mitochondrial single stranded binding protein (mtSSB), a core member of the mtDNA replisome and has a reduced set of mitochondrial transcription termination factors with only two of the four genes found in *Drosophila*.

Studies of human mt diseases have demonstrated that two classes of disorders will result in mt genome deletions causing minicircles similar to those found in *Pediculus* and *Damalinia*: 1) defects in nucleotide translocator proteins; and 2) defects in any of the mtDNA replicative enzymes [39-41]. Defects in the nucleotide translocator proteins cause imbalances in the intramitochondrial nucleotide pool and so result in a reduction in the rate and accuracy of mtDNA replication. Each of the three nucleotide translocator genes implicated in mtDNA deletions, ADP/ATP carrier protein (= adenine nucleotide translocator 1), deoxynucleoside kinase and ADP-forming succinyl-CoA synthetase, had homologues in *Pediculus* with very high to almost perfect levels of amino acid conservation across insects. There is thus no reason to suppose that the nucleotide translocators are functioning

abnormally in *Pediculus*, suggesting that nucleotide imbalances are not the cause of mtDNA minicircles. Of the four core enzymes responsible for mtDNA replication, the replicative helicase Twinkle and DNA polymerase  $\gamma$ A (POLGA) are similar to homologues in *Drosophila*, DNA polymerase  $\gamma$ B (POLGB) is quite divergent and mitochondrial-specific single-stranded binding protein (mtSSB) is apparently absent. Twinkle in *Pediculus* is moderately divergent from that found in *Drosophila* with limited amino-acid identity through the N-terminal primase domain; of the six conserved motifs only motifs 1 (zinc finger motif), 4 ( $\beta$ -sheet 10,  $\alpha$ -helix 3), 5 ( $\beta$ 12,  $\alpha$ 6), and 6 ( $\beta$ 13,  $\alpha$ 7) are conserved. There is much greater similarity in the central linker and C-terminal helicase domains; all of the five conserved motifs in the helicase domain are extremely conserved with the majority of residues identical and most of the remaining ones structurally similar (the structural model of Twinkle follows [19] and [42]). POLGA in *Pediculus* has a similar structure as in other metazoans; the N-terminal exonuclease (exo) domain has moderate similarity to other insects; the C-terminal polymerase (pol) and central linker domains are highly conserved and each of three active sites in both the exo and pol domains are highly conserved across insects (model following [43]). In contrast POLGB in *Pediculus* is highly divergent from that of other insects, lacking much of the N-terminal domain including the helix-loop-helix and leucine-zipper motifs, the central M-domain has limited similarity to other insects however the C-terminal domain is highly conserved (model following [44]). This suggests that a minimal mt replisome consisting of a DNA-polymerase (POLG holoenzyme) plus a helicase (Twinkle) is present in *Pediculus*, however, the absence of mtSSB is unique amongst sequenced insects.

## Discussion

### Mt genome rearrangements in Lice

As in all previous studies of lice [8,26,27] all of the mt genomes sequenced for the present study were highly rearranged relative to the inferred ancestral mt genome of insects. Mt genomes from different louse species are also highly rearranged relative to each other. One hundred and seventy two derived gene boundaries are found in the six complete and four minicircular mt genomes sequenced from lice to date (see Additional file 4 for complete list), dwarfing the 12 plesiomorphic gene boundaries which have been retained from the ancestral insect. This pattern is also reflected in every louse species examined where the numbers of derived boundaries is high compared to the number of plesiomorphic boundaries which range from a maximum of five plesiomorphic boundaries in *Bothriometopus* down to a minimum of none in *Ibidoecus*. This confirms earlier claims that lice possess the most rearranged mt genomes amongst the arthropods and are amongst the

most rearranged of all animals [26]. The 172 derived gene boundaries are made up of 161 unique gene boundaries found only in a single species, plus 11 shared derived boundaries which occur a total of 31 times across the examined louse species. These ratios are comparable to those found for rearrangements within Hymenoptera where of 67 rearrangements found in the order, 60 were unique to a single wasp species [9].

The present study is the first to find derived gene arrangements that are shared between distantly related louse species. The conserved rearrangements between the two goniodid genera, *Coloceras* and *Campanulotes* confirms the finding of Covacin et al. [27], based on partial genome sequences, that derived gene arrangements are shared at shallow phylogenetic levels within lice. More interestingly, we find the first evidence for shared, derived rearrangements which delimit much higher level groups within the lice e.g. *trnI-cox1* which appears to be synapomorphic for the Ischnocera although it has been lost in *Bothriometopus* and in the Anoplura; or *rrnL-trnW* and *trnM-rrnS* which define a clade consisting of the Goniodidae + (*Quadriceps*+*Philopterus*). Our finding of synapomorphic gene rearrangements within lice is in line with data from other insect orders including Hymenoptera [9,45], Neuroptera [46], Lepidoptera [47], Orthoptera [48,49] and Hemiptera [50], suggesting that rearrangements, when present, are often synapomorphic for major groups within each order. Lice however differ from all other insect orders in the almost complete absence of plesiomorphic gene boundaries as noted above. In contrast, all of the other insect orders cited above except Hymenoptera have only a single derived rearrangement with the remainder of the genome sharing the arrangement of the ancestral insect. Even the most rearranged hymenopteran mt genomes, those of *Bombus* and *Apis*, have only seven derived rearrangements with the rest of the genome in the ancestral arrangement [9,51].

### Minicircles in Lice

The present study shows that mitochondrial minicircles are common in lice being found in six of the ten louse species which have been studied to date. Furthermore, they are evolutionarily widespread within lice with representatives of all three ischnoceran families plus the Anoplura possessing minicircular mt genomes. Minicircles have yet to be reported from the other major louse sub-order, Amblycera. However little effort has yet been directed towards this group with just one genome sequenced (*Heterodoxus*) versus nine from Ischnocera. We have also identified two additional types of minicircles that could represent intermediate steps between the canonical single chromosome mt genomes found in almost all bilaterians and the small, multichromosomal



mt genomes found in sucking lice [16]. While it is tempting to see these two additional minicircles types as an evolutionary series leading to the multichromosomal type 3 minicircle, phylogenetic mapping of the minicircle types simply doesn't support such an interpretation (Figure 4a). The three minicircle types do not form a nested series within the tree and the plesiomorphic, full sized genome is found throughout the tree rather than being confined to some paraphyletic assemblage including the outgroup.

The type 1 minicircles, heteroplasmic genome reductions from a full sized mt genome, are confined to the Goniodidae but however likely represent the first step towards mt genome size reduction. This minicircle type is similar to the  $\Delta$ mtDNA state found widely in animals and similar mechanisms of DNA deletion could generate the heteroplasmic mt genome variants seen within the Goniodidae. As noted by Shao et al. [16], it is highly unlikely that the much more derived minicircle type seen in *Pediculus* evolved from a "big bang" which fragmented a full-sized mt genome into 18 minicircular chromosomes in a single event. A first intermediate step resembling the heteroplasmic mt genomes seen in *Coloceras* is much more likely even though it is clear that goniodids are not the ancestors of all other ischnoceran lice. Such heteroplasmic genome reductions, however, could not lead to the other minicircles types described here unless multiple independent deletions from the full sized mt genome resulted in a set of mt chromosomes which collectively covered all 37 mt genes, similar to the set of mt chromosomes found in *Globodera* [14,15]. Our available data suggests that there is just one  $\Delta$ mtDNA chromosome within the *Coloceras* specimens examined. It is also unknown whether the smaller mt genome type found in *Coloceras* is part of the germ line or if the deletion has occurred within the life span of the louse. Examination of additional specimens could shed light on how widely this particular minicircle is found within this species. Human, nematode and *Drosophila* examples of  $\Delta$ mtDNA, while not germline, are however, generated in a consistent fashion by deletions between repetitive regions [52-55], suggesting that testing between germline and age-related mt degeneration in goniodids will require targeted sequencing of eggs or early instar nymphs. In addition, further study is required to verify if louse species from which full sized genomes have been reported such as *Bothriometopus* and *Ibidoecus* do not also possess type 1 minicircles and if so what the relative balance of genomes types are within different species. The mt genome of *Bothriometopus* possesses multiple repetitive elements including both duplicated tRNAs and non-coding regions which could delimit deletion regions [8].

Type 2 minicircles, genomes with several full length major genes (PCGs or rRNA) but no evidence of a full

sized genome, would form the logical missing link between the heteroplasmic type 1 minicircles and the fully reduced type 3 minicircles. The type 2 minicircles that we found most closely resembled the minicircles reported from the nematode *Globodera* [14,15]; however, they differ in several respects. Unlike in *Globodera* we were unable to demonstrate multiple chromosomes. However, our failure to link other mt genes to the finished minicircles suggests the presence of other chromosomes. We also did not find evidence for a mosaic genome with the same gene being found on multiple chromosomes as slight sequence variants as was found in *Globodera*. Type 2 minicircles were also the most widespread minicircle type found in lice occurring in two different clades within the Ischnocera (Figure 4a) which would explain the difficulties which we have encountered trying to amplify louse mt genomes via the traditional long PCR approach. It is our prediction that further surveys will find type 2 minicircles to occur widely within the Ischnocera. However, experimental efforts must be made in each instance to exclude the possibility of apparently type 2 minicircles actually being type 1 minicircles i.e. that a full size mt genome is also present.

The type 3 minicircles (multichromosomal mt genomes with one major gene and a large, conserved non-coding region on each chromosome) are confined to the Anoplura and the Trichodectidae, which are now strongly supported on both morphological and molecular grounds to form a sister-group [31]. The difference between type 2 and type 3 minicircles is largely one of scale: several major genes are present in type 2 but not more than one major gene in type 3 minicircles; the number of tRNAs present in each type is variable but usually more are present in type 2 minicircles (three+ in each studied type 2 minicircle, not more than three in type 3 minicircles); and short non-coding regions without conserved sequence blocks in type 2 versus longer non-coding regions with several conserved sequence blocks in type 3. Additional studies may reveal that type 2 and type 3 minicircles are simply arbitrary points on a continuum of reduced mt chromosome sizes, but given available data, it is convenient to treat them separately at this time. The high degree of sequence conservation within the non-coding regions of each species is likely due to concerted evolution mediated by intramitochondrial recombination [56], and indeed Shao & Barker [32] report chimeric mitochondrial genomes in *Pediculus* composed of the fusion of two minicircles at the non-coding regions. The conserved sequence blocks mediate homologous recombination between different mitochondrial chromosomes [32]. The non-coding regions of *Pediculus* and *Damalinia* mt minicircles show a high degree of structural similarity. There are moderate differences in the sequence of

the non-coding region cloned from different chromosomes in each species, however, the entire non-coding region is readily alignable within each species. In both species, the non-coding region has three conserved sequence blocks located at the beginning, middle, and end of the non-coding region. There are also differences between non-coding regions from the two species. Unlike *Pediculus*, the non-coding regions of *Damalinia* mt minicircles are much smaller, ranging in length from 694 to 1098 bp vs 1643 to 2050 bp long in *Pediculus*. The location of a putative stem-loop structure differs in being located within conserved block 1 of *Damalinia* as contrasted to conserved block 2 of *Pediculus*. *Pediculus* lacks the macro-repeat regions that account for much of the variability between the non-conserved regions from different *Damalinia* chromosomes. Finally, there is very limited sequence conservation between the two species, even between the corresponding conserved regions. The other remarkable characteristic of the non-coding regions of these two species is that they are substantially larger than any of the non-coding regions reported from other louse mt genomes. These range in size from 48 bp in *Coloceras* to 284bp in *Bothriometopus*, whereas the smallest non-coding region in *Damalinia* is 694 bp from the *trnI-cox1* chromosome and in *Pediculus* is 1643 bp for one of the *nad5* chromosomes. The non-coding regions of *Damalinia* and *Pediculus* mt genomes are much closer in size to those found in many other insect orders which retain both the insect ancestral genome arrangements and the canonical single mt chromosome structure. The nine clones of non-coding regions from *Pediculus* minicircles that were fully sequenced are all larger than the coding regions on their respective chromosomes [16]. While the size of the non-coding region has not been determined for all 18 mt chromosomes identified in *Pediculus*, the sequence conservation between non-coding regions suggests that they will all be 1600bp plus in size and potentially many times larger than many of the coding regions, which were as small as 65bp for one chromosome that contains only a single tRNA gene. It seems counter intuitive that whatever evolutionary forces may have been driving lice towards smaller, minicircular mt genomes has also driven a large expansion in the size of the non-coding regions. Many additional mt genomes from this group will need to be investigated to determine if there truly is a trend toward size increases in the non-coding regions of lice with type 3 minicircular mt genomes and what, if any, function these bloated non-coding regions serve in genome maintenance, replication, or transcription.

#### Factors Leading to Mitochondrial Minicircles in Lice

Given that minicircularized mitochondrial chromosomes are very rare across the diversity of life, it is interesting

to consider what factors may have given rise to their evolution. Shao et al. [16] proposed that the evolution of minicircular mt genomes coevolved with blood-feeding in the sucking lice (Anoplura). However this hypothesis is contradicted by the available mt genome data. Minicircular mt genomes within lice are not confined to blood-feeding taxa being also found in four of the seven philopterid genera studied which are all feather-feeding [57] and in the trichodectid *Damalinia* which is skin-feeding [58]. Furthermore, minicircular mt genomes are not found in other blood-feeding lice such as the boopiid *Heterodoxus* [26], or in other blood-feeding arthropods such as mosquitoes [59], reduviid bugs [60], or ticks [61], or in other blood-feeding animals such as leeches (Wu et al. *unpublished data*). Within insects, minicircular mt genomes are confined to lice and the specific minicircular mt genome type documented in [16] is confined to a clade minimally consisting of Anoplura + Trichodectidae [31]. Rather than an extrinsic, life-history related explanation relating minicircles in lice to their parasitism, an explanation based on a genetic factor, loss of mtSSB, seems more plausible.

Of the subset of genes which have been related to mt genome deletions, annotation of the *Pediculus* genome found that the only one that was absent, or highly divergent from homologues in other insects, was mtSSB. While it is possible that mtSSB is present in *Pediculus* but so highly modified that our BLAST approach to gene identification did not detect it, the high degree of conservation of this gene in other insects (Additional File 5) suggests that even if it were present it is so highly modified in *Pediculus* as to be non-functional in which case an argument based on its absence is still valid (c.f. [62]).

Functionally mtSSB has five major roles in the upkeep of mt genomes:

1. Binding to single-stranded mtDNA during replication, protecting it from mutation [63,64].
2. Preventing DNA-slippage at stalled replication forks [65].
3. Increasing the processivity of the replisome thus increasing the rate of DNA synthesis [66,67].
4. Stabilizing the triple-stranded D-loop replicative intermediate in vertebrates [68].
5. A major component of mitochondrial nucleoids [64,69,70].

All single-stranded binding proteins have a function in protecting DNA during the single stranded phase of replication [63] and the absence of mtSSB may be responsible for heightened rates of substitution observed within lice [71,72]. The second and third of these functions suggest that in the absence of mtSSB, formation of

minicircles would be favored and possibly even evolutionarily adaptive. Viguera et al. [65] have demonstrated that re-initiation of stalled replication forks requires dissociation of the polymerase from the template. In such circumstances, the single-stranded nascent DNA strand can reassociate with other regions of the template strand resulting in replication slippage as the intervening region between the original and new template sites is not replicated. SSBs stabilize the nascent strand while the polymerases are detached and prevent the illicit pairings which result in slippage in a dose dependent manner. That is, low concentrations of SSB result in a mixed population of full length and partial genomes whereas high concentrations can prevent the formation of slipped genome altogether [65]. This model of slippage depends on the presence of repetitive DNA stretches such that the nascent strand could anneal to both locations. This is enhanced by the formation of hairpin-loop secondary structures that bring the repetitive regions into close physical proximity and can cause replication to stall within the repeat region. Just such a deletion between repetitive elements was identified in the *Coloceras* mt genome resulting in the heteroplasmic mix of full length and minicircular mt genomes.

The absence of mtSSB would also have favoured the retention of minicircles once formed. mtSSB stabilizes the single-stranded conformation of the lagging replication strand which increases the rate of replication (processivity of the replisome) and the maximum size of templates which can be replicated. Korhonen et al. [66] have demonstrated that mtSSB increases the size of replication products in a dose dependent manner from around 2000 bp in its absence to over 15000 bp at high concentrations. In the absence of mtSSB complete replication of a full sized mt genome may be impossible. However, the minicircle sized genomes might be sufficiently small to be replicated by polymerase and helicase alone. The minicircles found in *Pediculus* and *Damalinia* are 3-4 and 2-2.5 kb in size respectively, comparable to the replicable sizes found by experimental knock out of mtSSB. It is thus possible to propose that loss of mtSSB would favour both the formation of minicircles through replicative slippage and their intergenerational retention due to an impaired ability to replicate larger mt genomes.

The effect of loss of mtSSB on the final two roles - D-loop stabilization and nucleoid formation - are difficult to assess owing to a complete lack of research on these phenomena in insect models. In mammals mtSSB stabilizes D-loops by reducing the capacity of transcription factor A (TFAM) to remove the RNA priming region and resolve triple-stranded, DNA/RNA hybrid D-loops into double-stranded non-replicative DNA [68]. D-loops have not been directly demonstrated within insects [73];

however, the conserved sequence box (CSB) regions which are critical to D-loop formation in mammals have [74,75]. CSB regions are binding sites for transcription termination factors and the premature termination of transcription at these sites generates the short RNA primers that form the D-loop [76]. It is unknown what, if any, interactions between mtSSB and TFAM are responsible for D-loop persistence in insects and accordingly the impact of the loss of mtSSB on replication initiation cannot be predicted. TFAM has also been shown in mammals to control mtDNA copy number, possibly due to its role in nucleoid formation [77]. TFAM, however, can be excluded as a cause of minicircle formation in *Pediculus* as its structure is very similar to that of TFAMs from vertebrates and other insects in that it possesses two conserved high mobility group (HMG) motifs (the third, C-terminal HMG found in *Drosophila* and *Anopheles* is however absent but is probably Diptera specific). Mitochondrial nucleoids are one of the most poorly understood facets of mtDNA maintenance, composed of multiple mtDNA molecules complexed with protein and have been proposed as the units of mt inheritance [78]. Nucleoids have, however, yet to be studied in an insect model. Most available data on nucleoid composition is from yeast or vertebrate model species and the protein complement of the nucleoid varies widely between different species [66,70,79]. mtSSB is one of the few consistently recovered components of the vertebrate mt nucleoid; however, its function within the nucleoid is unknown and accordingly it is unclear what effect its absence would have.

## Conclusion

Genome deletions and minicircles are common in louse evolution, not being confined to any particular louse subgroup. Three types of minicircles, which do not appear to form an evolutionary series, were identified within lice, however Type 3 minicircles (multichromosomal minicircles with few genes) are likely synapomorphic for the group Trichodectidae+Anoplura. Annotation of the ne-mt genes from the *Pediculus* nuclear genome, suggests that mitochondrial function is largely intact, as the majority of genes are present, despite the extreme divergence of its multichromosomal mt genome from the single chromosomal mt genome found in almost all other bilaterian animals. The absence of one ne-mt gene, mtSSB, could explain the fragmentation of the mt genome into multiple chromosomes. This mechanistic explanation for the occurrence of minicircles in lice is more likely than ones based on a blood-feeding life-history (see [16]). The absence of mtSSB may also have favoured the retention of minicircles once formed due to an incapacity of the mt replisome to replicate full sized mt genomes. The effect of

lacking mtSSB would have on other mt functions such as D-loop stabilization and nucleoid formation requires additional research.

## Methods

### PCR amplification and sequencing

Collecting records for louse specimens are given in Additional File 1. All specimens were collected into 100% ethanol and stored at -80°C prior to extraction. DNA was extracted using the DNEasy Tissue Kit (Qia-gen, Hilden, Germany) from a single louse and a single DNA extract was used for all PCRs in this study. Sequence and location of all amplification primers used in this study are in Additional file 6. Initial PCRs for *Ibidoecus* amplified three intragenic regions: 383 bp of *cox1* (primers: L6225/H7005 from [80]); 550 bp of *rrnL* (16Sar/16Sbr from [81]); and 350 bp of *rrnS* (12Sai/12Sbi from [81]) amplified using reaction conditions as described in [82]. These initial sequences were used to design specific long PCR primers which were combined in all possible forward and reverse combinations of attempted long PCRs between the three genes (standard long PCR protocol below). Initial PCRs for *Coloceras*, *Anaticola*, *Philopterus* and *Quadriceps* consisted of two long PCRs *rrnL* to *rrnS* (CAM10/CAM7 from [27]) and *rrnS* to *rrnL* (GON1/GON2 from [27]). Long PCRs for bird lice were performed using Elongase (Invitrogen, Carlsbad, California) with the following cycling conditions: 92°C for 2 min; 40 cycles of 92°C for 30 sec, 50°C for 30 sec, 68°C for 12 min; and a final extension step of 68°C for 20 min. Sequencing was performed using ABI BigDye ver. 3 dye terminator sequencing technology and run on ABI 3770 or ABI 3740 capillary sequencers. Cycle sequencing conditions were 28 cycles of 94°C/10 sec, 50°C/5 sec, 60°C/4 min. Within each long PCR product the full, double stranded sequence was determined by primer walking (primer sequences available from SLC on request). Initial PCRs for *Damalinia* were generated for *cox1* (L6225/H7005), *rrnL* (16Sar/16Sbr) and *rrnS* (12Sai/12Sbi) using ExTaq (TaKaRa, Otsu, Shiga) with the following cycling conditions: 94°C for 2 min, 40 cycles of 94°C for 30 sec, 45°C for 30 sec, 65°C for 5 min. Long PCRs for *Damalinia* were generated using LA-Taq (TaKaRa, Otsu, Shiga) with the following cycling conditions: 98°C for 2 min, 40 cycles of 98°C for 30 sec, and 68°C for 15 min (combined annealing/extension step). Due to sequence variability, the non-coding regions of *Damalinia* were cloned prior to sequencing using the pGem-T Easy Vector system (Promega, Madison, Wisconsin) following manufacturer protocols. Sequencing of short PCR, long PCR and cloned fragments from *Damalinia* was performed using the CEQ DNA analysis system (Beckman Coulter, Brea, California) following manufacturer protocols.

### Analysis and Annotation

Raw sequence files were proof read and assembled into contigs in Sequencher ver. 4 (GeneCodes Corporation, Ann Arbor, Michigan). Transfer RNA analysis was conducted using tRNAscan-SE [83] with invertebrate mitochondrial codon predictors and a cove score cut off of 1. Reading frames between tRNAs were found in Sequencher and identified using translated BLAST searches (blastx) [84] as implemented at the NCBI website (<http://www.ncbi.nlm.nih.gov/>). Annotations of ribosomal RNA gene boundaries were performed by comparison with published models of rRNA secondary structure (e.g. [85]). Evolution of genome structure and rearrangements was assessed by comparison with phylogenetic trees inferred by adding sequences for *cox1*, *lsu-rRNA* and *wingless* to the alignment used in [31] for the five species included in the present study but not included, *Ibidoecus*, *Bothriometopus*, *Coloceras*, *Quadriceps* and *Philopterus*. Trees were inferred using MrBayes ver3.1.1 [86] using two independent runs, each with 4 chains and unlinked data partitions for 10,000,000 generations and trees were sampled every 1000 generations. The first 1000 trees were discarded as burnin following determination of stationarity in the Bayesian analyses using Tracer ver. 1.4 [87]. Substitution models were estimated for each gene partition using AIC as implemented in MrModeltest 2.3 [88]. Genome rearrangements were coded as binary states or unknown for species from which only partial mtgenomes are available. Character evolution was traced using MacClade ver 4.06 [89] treating character states as unordered. The tree has been uploaded to TreeBASE (accession: <http://purl.org/phylo/treebase/phyloids/study/TB2:S11666>)

Annotation of the nuclear-encoded, mitochondria-targeted (ne-mt) genes within the *Pediculus humanus* genome were conducted using genome assembly phumU1, hosted on the VectorBase website [90]. The most up to date list of ne-mt genes identified from *Drosophila* was compiled from [38,91]. *Pediculus* orthologues of the *Drosophila* ne-mt genes were identified by the best bidirectional hit (BBH) approach [92]. Amino acid sequences for the *Drosophila melanogaster* copies of each gene were taken from GenBank and blasted against the predicted peptide set for *Pediculus* (Phum\_pep\_ver1.1 gene build) using blastP [84] as implemented on VectorBase. Best blast matches within the *Pediculus* genome were confirmed by reciprocal blasting of the inferred peptide against the *Drosophila* genome on GenBank. Only reciprocal best blast pairs are considered orthologous genes and treated further within this study.

### Additional material

**Additional file 1: PCR strategy.** The PCR amplification strategy and collection records for the lice sequenced in this study.



**Additional file 2: Full phylogenetic tree of lice and relatives.** Full tree generated after [31] which was pruned to produce figure 2.

**Additional file 3: ne-mt genes found in *Pediculus*.** Complete list of gene homologues and blast statistics for the annotation of ne-mt genes found in *Pediculus*.

**Additional file 4: Novel mt gene boundaries in lice.** Complete list of novel mt gene boundaries found in lice.

**Additional file 5: mtSSB alignment.** Alignment of the amino acid sequences of mtSSB genes annotated from insect nuclear genomes.

**Additional file 6: Primers.** Primer sequences and combinations used in this study.

## Acknowledgements

This study was supported by the NSF grants DEB0444972 (SLC, MFW, KPJ) and DEB0612938 (KPJ), CSIRO's Julius Career Award (SLC), JSPS Grant Numbers 13740486, 18770058 and 21770083 (KY) and the Global COE Program Establishment of the Center for Integrated Field Environmental Science at Hokkaido University (KY, AM). The authors also thank Koichi Kaji, Kevin McCracken, Hiroshi Takahashi, Jason Weckstein, and Robert Wilson for their assistance in field work.

## Author details

<sup>1</sup>Discipline of Biogeosciences, Faculty of Science & Technology, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia.

<sup>2</sup>Australian National Insect Collection & CSIRO Ecosystem Sciences, Black Mountain Laboratories, PO Box 1700, Canberra, ACT 2601, Australia.

<sup>3</sup>Systematic Entomology, Graduate School of Agriculture, Hokkaido University, Sapporo, 060-8589, Japan. <sup>4</sup>Department of Biology, Brigham Young University, Provo, UT, 84602, USA. <sup>5</sup>Illinois Natural History Survey, University of Illinois, 1816 South Oak St., Champaign, IL, 61820, USA.

## Authors' contributions

Conceived this study: SLC & KPJ. Performed the work and statistical analyses: SLC, KY, AM, MFW & KPJ. Wrote the paper: SLC, KY, MFW & KPJ. All authors read and approved the final manuscript.

Received: 23 February 2011 Accepted: 4 August 2011

Published: 4 August 2011

## References

- Lavrov DV, Brown WM: *Trichinella spiralis* mtDNA: A nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate metazoans. *Genetics* 2001, **157**:621-637.
- Domes K, Maraun M, Scheu S, Cameron SL: The complete mitochondrial genome of the sexual oribatid mite *Steganacarus magnus*: genome rearrangements and loss of tRNAs. *BMC Genomics* 2008, **9**:532.
- Mallatt J, Craig CW, Yoder MJ: Nearly complete rRNA genes assembled from across the metazoan animals: Effects of more taxa, a structure-based alignment and paired-sites evolutionary models on phylogeny reconstruction. *Mol Phylogenet Evol* 2010, **55**:1-17.
- Boore JL: Animal mitochondrial genomes. *Nucleic Acids Res* 1999, **27**:1767-1780.
- Webster BL, Mackenzie-Dodds JA, Telford MJ, Littlewood DTJ: The mitochondrial genome of *Priapulus caudatus* Lamarck (Priapulida: Priapulidae). *Gene* 2007, **389**:96-105.
- Braband A, Cameron SL, Podsiadlowski L, Daniels SR, Mayer G: The mitochondrial genome of the onychophoran *Opisthopterus cinctipes* (Peripatopsidae) reflects the ancestral mitochondrial gene arrangement of Panarthropoda and Ecdysozoa. *Mol Phylogenet Evol* 2010, **57**:285-295.
- Braband A, Podsiadlowski L, Cameron SL, Daniels SR, Mayer G: Extensive duplication events account for multiple control regions and pseudogenes in the mitochondrial genome of the velvet worm *Metaperipatus inae* (Onychophora, Peripatopsidae). *Mol Phylogenet Evol* 2010, **57**:293-300.
- Cameron SL, Johnson KP, Whiting MF: The mitochondrial genome of the screamer louse *Bothriometopus* (Phthiraptera: Ischnocera): Effects of extensive gene rearrangements on the evolution of the genome as a whole. *J Mol Evol* 2007, **65**:589-604.
- Dowton M, Cameron SL, Dowd J, Austin AD, Whiting MF: Characterisation of 67 mitochondrial gene rearrangements in the Hymenoptera reveal underlying trends in mitochondrial genome evolution. *Mol Biol Evol* 2009, **26**:1607-1617.
- Mueller RL, Boore JL: Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. *Mol Biol Evol* 2005, **22**:2104-2112.
- Gray MW, Lang BF, Burger C: Mitochondria of protists. *Annu Rev Genet* 2004, **38**:477-524.
- Cummings DJ: Mitochondrial genomes of the ciliates. *Int Rev Cytol* 1992, **141**:1-64.
- Burger G, Gray MW, Lang BF: Mitochondrial genomes: anything goes. *T Genet* 2003, **19**:709-716.
- Gibson T, Blok VC, Phillips MS, Hong G, Kumarasinghe D, Riley IT, Dowton M: The mitochondrial subgenomes of the nematode *Globodera pallida* are mosaics: Evidence of recombination in an animal mitochondrial genome. *J Mol Evol* 2007, **64**:463-471.
- Gibson T, Blok VC, Dowton M: Sequence and characterisation of six mitochondrial subgenomes from *Globodera rostochiensis*: Multipartite structure is conserved among close nematode relatives. *J Mol Evol* 2007, **65**:308-315.
- Shao R, Kirkness EF, Barker SC: The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. *Genome Res* 2009, **19**:904-912.
- Holt IJ: Zen and the art of mitochondrial DNA maintenance. *Trends Genet* 2010, **26**:103-109.
- Geromel V, Kadhon N, Cebalos-Picot I, Ouari O, Polidori A, Munnich A, Rötig A, Rustin P: Superoxide-induced massive apoptosis in cultured skin fibroblasts harboring the neurogenic ataxia retinitis pigmentosa (NARP) mutation in the ATPase-6 gene of the mitochondrial DNA. *Hum Mol Genet* 2001, **10**:1221-1228.
- Spelbrink JN, Li FY, Tiranti V, Nikali K, Yuan QP, Tariq M, Wanrooij S, Garrido N, Comi G, Morandi L, Santoro L, Toscano A, Fabrizi GM, Somer H, Crozen R, Beeson D, Poulton J, Suomalainen A, Jacobs HT, Zeviani M, Larsson C: Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. *Nat Genet* 2001, **28**:223-231.
- van Goethem G, Dermaut B, Löfgren A, Martin JJ, Broeckhoven C: Mutation of *POLG* is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. *Nat Genet* 2001, **28**:211-212.
- Saada A: Deoxyribonucleotides and Disorders of Mitochondrial DNA Integrity. *DNA Cell Biol* 2004, **23**:797-806.
- Fleming JE, Miquel J, Cottrell SF, Yengoyan LS, Economos AC: Is cell aging caused by respiration-dependent injury to the mitochondrial genome? *Gerontology* 1982, **28**:44-53.
- Pak JW, Herbst A, Bua E, Gokey N, McKenzie D, Aiken JM: Mitochondrial DNA mutations as a fundamental mechanism in physiological declines associated with aging. *Aging Cell* 2003, **2**:1-7.
- Jacobs HT: The mitochondrial theory of aging: dead or alive. *Aging Cell* 2003, **2**:11-17.
- Sánchez-Martínez Á, Luo N, Clemente P, Adán C, Hernández-Sierra R, Ochoa P, Fernández-Moreno MÁ, Kaguni LS, Garesse R: Modelling human mitochondrial diseases in flies. *Biochim Biophys Acta* 2006, **1757**:1190-1198.
- Shao R, Campbell NJH, Barker SC: Numerous gene rearrangements in the mitochondrial genome of wallaby louse, *Heterodoxus macropus* (Phthiraptera). *Mol Biol Evol* 2001, **18**:858-865.
- Covacin C, Shao R, Cameron SL, Barker SC: Extraordinary amounts of gene rearrangement in the mitochondrial genomes of lice (Insecta: Phthiraptera). *Insect Mol Biol* 2006, **15**:63-68.
- Johnson KP, Whiting MF: Multiple genes and the monophyly of Ischnocera (Insecta: Phthiraptera). *Mol Phylogenet Evol* 2002, **22**:101-110.
- Johnson KP, Yoshizawa K, Smith VS: Multiple origins of parasitism in lice. *P Roy Soc B - Biol Sci* 2004, **271**:1771-1776.
- Murrell A, Barker SC: Multiple origins of parasitism in lice: phylogenetic analysis of SSU rDNA indicates that the Phthiraptera and Psocoptera are not monophyletic. *Parasitol Res* 2005, **97**:274-280.
- Yoshizawa K, Johnson KP: How stable is the "Polyphyly of Lice" (Insecta: Phthiraptera) hypothesis? A comparison of phylogenetic signal in multiple genes. *Mol Phylogenet Evol* 2010, **55**:939-951.



32. Shao R, Barker SC: Chimeric mitochondrial minichromosomes of the human body louse, *Pediculus humanus*: Evidence for homologous and non-homologous recombination. *Gene* 2011, **473**:36-43.
33. Kirkness EF, et al: Genome sequences of the human body louse and its primary endosymbiont: Insights into the permanent parasitic lifestyle. *P Natl Acad Sci USA* 2010, **107**:12168-12173.
34. Tripoli G, D'Elia D, Barsanti P, Caggese C: Comparison of the oxidative phosphorylation (OXPHOS) nuclear genes in the genomes of *Drosophila melanogaster*, *Drosophila pseudoobscura* and *Anopheles gambiae*. *Genome Biol* 2005, **6**:R11.
35. D'Elia D, Catalano D, Licciulli F, Turi A, Tripoli G, Porcelli D, Saccone C, Caggese C: The MitoDrome database annotates and compares the OXPHOS nuclear genes of *Drosophila melanogaster*, *Drosophila pseudoobscura* and *Anopheles gambiae*. *Mitochondrion* 2006, **6**:252-257.
36. Saccone C, Lanave C, De Grassi A: Metazoan OXPHOS gene families: Evolutionary forces at the level of mitochondrial and nuclear genomes. *Biochim Biophys Acta* 2006, **1757**:1171-1178.
37. Porcelli D, Barsanti P, Pesole G, Caggese C: The nuclear OXPHOS genes in Insecta: a common evolutionary origin, a common cis-regulatory motif, a common destiny for gene duplicates. *BMC Genomics* 2007, **7**:215.
38. Smits P, Smeitink JAM, van den Heuvel LP, Huynen MA, Ettema TJG: Reconstructing the evolution of the mitochondrial ribosomal proteome. *Nucleic Acids Res* 2007, **35**:4686-4703.
39. Wanrooij S, Luoma P, van Goethem G, van Broeckhoven C, Suomalainen A, Spelbrink JN: Twinkle and POLG defects enhance age-dependent accumulation of mutations in the control region of mtDNA. *Nucleic Acids Res* 2004, **32**:3053-3064.
40. Alberio S, Mineri R, Tiranti V, Zeviani M: Depletion of mtDNA: Syndromes and genes. *Mitochondrion* 2007, **7**:6-12.
41. Ashley N, Adams S, Slama A, Zeviani M, Suomalainen A, Andreu AL, Naviaux RK, Poulton J: Defects in maintenance of mitochondrial DNA are associated with intramitochondrial nucleotide imbalances. *Hum Mol Genet* 2007, **16**:1400-1411.
42. Shutt TE, Gray MW: Twinkle, the mitochondrial replicative DNA helicase, is widespread in the eukaryotic radiation and may also be the mitochondrial DNA primase in most eukaryotes. *J Mol Evol* 2006, **62**:588-599.
43. Kaguni LS: DNA polymerase  $\gamma$ , the mitochondrial replicase. *Annu Rev Biochem* 2004, **73**:293-320.
44. Fan L, Kaguni LS: Multiple regions of subunit interaction in *Drosophila* mitochondrial DNA polymerase: Three functional domains in the accessory subunit. *Biochemistry-US* 2001, **40**:4780-4791.
45. Cameron SL, Dowton M, Castro LR, Ruben K, Whiting MF, Austin AD, Diemert K, Stevens J: Mitochondrial genome organization and phylogeny of two vespid wasps. *Genome* 2008, **51**:800-808.
46. Cameron SL, Sullivan J, Song H, Miller KB, Whiting MF: A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. *Zool Scr* 2009, **38**:575-590.
47. Taylor MFJ, McKechnie SW, Pierce NE, Kreitman M: The lepidopteran mitochondrial control region: Structure and evolution. *Mol Biol Evol* 1993, **10**:1259-1272.
48. Flook PK, Rowell CHF, Gellissen G: Homoplastic rearrangements of insect mitochondrial tRNA genes. *Naturwissenschaften* 1995, **82**:336-337.
49. Fenn JD, Song H, Cameron SL, Whiting MF: A mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. *Mol Phylogen Evol* 2008, **49**:59-68.
50. Thao ML, Baumann L, Baumann P: Organization of the mitochondrial genomes of whiteflies, aphids and psyllids (Hemiptera: Sternorrhyncha). *BMC Evol Biol* 2004, **4**:25.
51. Crozier RH, Crozier YC: The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organisation. *Genetics* 1993, **133**:97-117.
52. Melov S, Lithgow GJ, Fischer DR, Tedesco PM, Johnson TE: Increased frequency of deletions in the mitochondrial genome with age of *Caenorhabditis elegans*. *Nucleic Acids Res* 1995, **23**:1419-1425.
53. Samuels DC: Mitochondrial DNA repeats constrain the life span of mammals. *Trends Genet* 2004, **20**:226-229.
54. Yui R, Matsuura ET: Detection of deletions flanked by short direct repeats in mitochondrial DNA of aging *Drosophila*. *Mutat Res* 2006, **594**:155-161.
55. Guo X, Popadin KY, Markuzon N, Orlov YL, Kraytsberg Y, Krishnan KJ, Zsurka G, Turnbull DM, Kunz WS, Khrapko K: Repeats, longevity and the sources of mtDNA deletions: Evidence from 'deletional spectra'. *Trends Genet* 2010, **26**:340-343.
56. Dowton M, Campbell NJH: Intramitochondrial recombination - is this why some mitochondrial genes sleep around? *Trends Ecol Evol* 2001, **16**:269-271.
57. Johnson KP, Clayton DH: The biology, ecology and evolution of chewing lice. In *The chewing lice: World Checklist and Biological Overview. Volume 24*. Edited by: Price RD, Hellenthal RA, Palma RL, Johnson KP, Clayton DH. Champaign: Illinois Natural History Survey Special Publication; 2003:449-476.
58. Sinclair AN, Butler RW, Picton J: Feeding of the chewing louse *Damalinia ovis* (Shrank) (Phthiraptera: Trichodectidae) on sheep. *Vet Parasitol* 1989, **30**:233-251.
59. Beard CB, Hamm DM, Collins FH: The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organisation and comparisons with mitochondrial sequences of other insects. *Insect Mol Biol* 1993, **2**:103-124.
60. Dotson EM, Beard CB: Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. *Insect Mol Biol* 2001, **10**:205-215.
61. Black WC, Roehrdanz RL: Mitochondrial genome order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. *Mol Biol Evol* 1998, **15**:1772-1785.
62. Huynen MA, Snel B, van Noort V: Comparative genomics for reliable protein-function prediction from genomic data. *Trends Genet* 2004, **20**:340-344.
63. Pestryakov PE, Lavrik OI: Mechanisms of single-stranded DNA binding protein functioning in cellular DNA metabolism. *Biochemistry (Mosc)* 2008, **73**:1388-1404.
64. Ruhanen H, Borrie S, Szabadkai G, Tyynismaa H, Jones AWE, Kang D, Taanman JW, Yasukawa T: Mitochondrial single-stranded DNA binding protein is required for maintenance of mitochondrial DNA and 7S DNA but is not required for mitochondrial nucleoid organization. *Biochim Biophys Acta* 2010, **1803**:931-939.
65. Viguera E, Canceill D, Ehrlich SD: Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J* 2001, **20**:2587-2595.
66. Korhonen JA, Pham XH, Pellegrini M, Falkenberg M: Reconstitution of a minimal mtDNA replisome *in vitro*. *EMBO J* 2004, **23**:2423-2429.
67. Falkenberg M, Larsson NG, Gustafsson CM: DNA replication and transcription in mammalian mitochondria. *Annu Rev Biochem* 2007, **76**:679-699.
68. Takamatsu C, Umeda S, Ohsato T, Ohno T, Abe Y, Fukuoka A, Shinagawa H, Hamasaki N, Kang D: Regulation of mitochondrial D-loops by transcription factor A and single-stranded DNA-binding protein. *EMBO Rep* 2002, **3**:451-456.
69. Bogenhagen DF, Wang Y, Shen EL, Kobayashi R: Protein components of mitochondrial DNA nucleoids in higher eukaryotes. *Mol Cell Proteomics* 2003, **2**:1205-1216.
70. Wang Y, Bogenhagen DF: Human mitochondrial DNA nucleoids are linked to protein folding machinery and metabolic enzymes at the mitochondrial inner membrane. *J Biol Chem* 2006, **281**:25791-25802.
71. Johnson KP, Adams RJ, Page RDM, Clayton DH: When do parasites fail to speciate in response to host speciation? *Syst Biol* 2003, **52**:37-47.
72. Yoshizawa K, Johnson KP: Phylogenetic position of Phthiraptera (Insecta: Paraneoptera) and elevated rate of evolution in mitochondrial 12S and 16S rDNA. *Mol Phylogen Evol* 2003, **29**:102-114.
73. Garesse R, Kaguni LS: A *Drosophila* model of mitochondrial DNA replication: Proteins, genes and regulation. *IUBMB Life* 2005, **57**:555-561.
74. Lewis DL, Farr CL, Farquhar AL, Kaguni LS: Sequence, organisation and evolution of the A+T rich region of *Drosophila melanogaster* mitochondrial DNA. *Mol Biol Evol* 1994, **11**:523-538.
75. Tsujino F, Kosemura A, Inohira K, Hara T, Otsuka YF, Obara MK, Matsuura ET: Evolution of the A+T rich region of mitochondrial DNA in the *melanogaster* species subgroup of *Drosophila*. *J Mol Evol* 2002, **55**:573-583.
76. Pham XH, Farge G, Shi Y, Gaspari M, Gustafsson CM, Falkenberg M: Conserved sequence box II directs transcription termination and primer formation in mitochondria. *J Biol Chem* 2006, **281**:24647-24652.
77. Ekstrand MI, Falkenberg M, Rantanen A, Park CB, Gaspari M, Hultenby K, Pustyn P, Gustafsson CM, Larsson NL: Mitochondrial transcription factor A

- regulates mtDNA copy number in mammals. *Hum Mol Genet* 2004, **13**:935-944.
78. Jacobs HT, Lehtinen SK, Spelbrink JN: **No sex please, we're mitochondria: A hypothesis on the somatic unit of inheritance of mammalian mtDNA.** *Bioessays* 2000, **22**:564-572.
  79. Garrido N, Griparic L, Kokitalo E, Wartiovaara J, van der Bliek AM, Spelbrink JN: **Composition and dynamics of human mitochondrial nucleoids.** *Mol Biol Cell* 2003, **14**:1583-1596.
  80. Johnson KP, Cruikshank RH, Adams RJ, Smith VS, Page RDM, Clayton DH: **Dramatically elevated rates of mitochondrial substitution in lice (Insecta: Phthiraptera).** *Mol Phylogen Evol* 2003, **26**:231-242.
  81. Simon C, Frati F, Beckenbach AT, Crespi B, Liu B, Flook P: **Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers.** *Ann Entomol Soc Am* 1994, **87**:651-701.
  82. Johnson KP, Adams RJ, Clayton DH: **Molecular systematics of Goniodidae (Insecta: Phthiraptera).** *J Parasitol* 2001, **87**:862-869.
  83. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res* 1997, **25**:955-964.
  84. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
  85. Cameron SL, Whiting MF: **The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths.** *Gene* 2008, **408**:112-123.
  86. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
  87. **Tracer v1.4.** [<http://beast.bio.ed.ac.uk/Tracer>].
  88. Nylander JAA: **MrModeltest v2. Program distributed by the authors** Evolutionary Biology Centre, Uppsala University; 2004.
  89. Maddison W, Maddison D: **MacClade ver 4.06.** Sunderland: Sinauer Associates; 2003.
  90. Lawson D, Arensburg P, Atkinson P, Besansky NJ, Bruggner RV, Butler R, Campbell KS, Christophides GK, Christley S, Dialynas E, Hammond M, Hill CA, Konopinski N, Lobo NF, MacCallum RM, Madey G, Megy K, Meyer J, Redmond S, Severson DW, Stinson EO, Topalis P, Birney E, Gelbart WM, Kafatos FC, Louis C, Collins FH: **VectorBase: a data resource for invertebrate vector genomics.** *Nucleic Acids Res* 2009, **37**:D583-D587.
  91. Sardiello M, Licciulli F, Catalano D, Attimonelli M, Caggese C: **MitoDrome: A database of *Drosophila melanogaster* nuclear genes encoding proteins targeted to the mitochondrion.** *Nucleic Acids Res* 2003, **31**:322-324.
  92. Hulsen T, Huynen MA, de Vlieg J, Groenen PMA: **Benchmarking ortholog identification methods using functional genomics data.** *Genome Biol* 2006, **7**:R31.

doi:10.1186/1471-2164-12-394

**Cite this article as:** Cameron *et al.*: Mitochondrial genome deletions and minicircles are common in lice (Insecta: Phthiraptera). *BMC Genomics* 2011 **12**:394.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

